

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
24 December 2003 (24.12.2003)

PCT

(10) International Publication Number
WO 2003/106672 A3

(51) International Patent Classification⁷: **C12N 15/00**

DNAFORM, 3-35, Mita 1-chome, Minato-ku, Tokyo 108-0073 (JP).

(21) International Application Number:
PCT/JP2003/007514

(74) Agents: **OKUYAMA, Shoichi** et al.; 8th Floor, Akasaka NOA Bldg., 2-12, Akasaka 3-chome, Minato-ku, Tokyo 107-0052 (JP).

(22) International Filing Date: 12 June 2003 (12.06.2003)

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(26) Publication Language: English

(30) Priority Data:
2002-171851 12 June 2002 (12.06.2002) JP
2002-235294 12 August 2002 (12.08.2002) JP

(71) Applicants (*for all designated States except US*): **RIKEN** [JP/JP]; 2-1, Hirosawa, Wako-shi, Saitama 351-0198 (JP). **KABUSHIKI KAISHA DNAFORM** [JP/JP]; 3-35, Mita 1-chome, Minato-ku, Tokyo 108-0073 (JP).

(72) Inventors; and

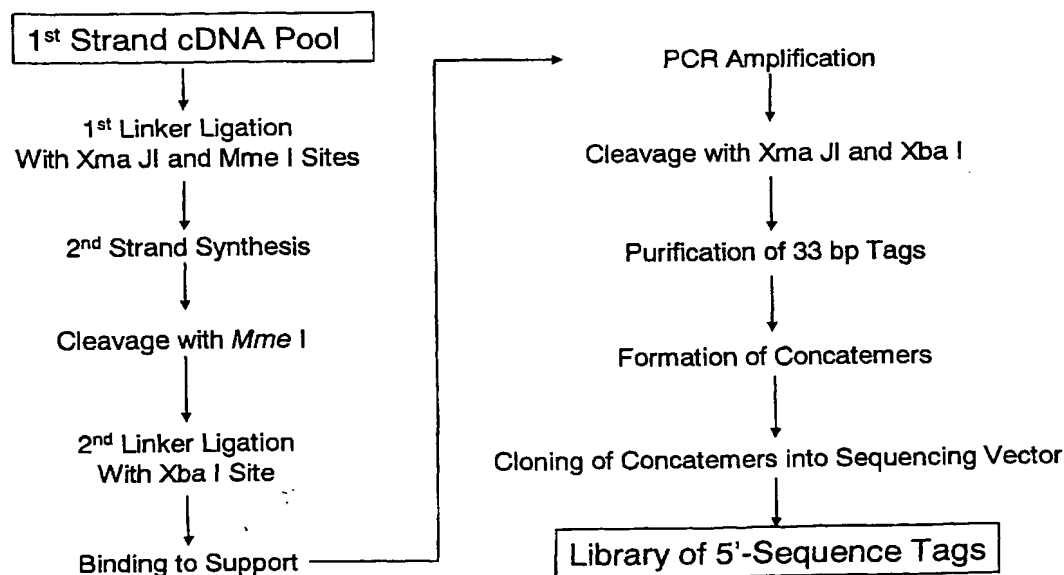
(75) Inventors/Applicants (*for US only*): **HAYASHIZAKI, Yoshihide** [JP/JP]; 22-8, Inarimae, Tsukuba-shi, Ibaraki 305-0061 (JP). **CARNINCI, Piero** [IT/JP]; c/o RIKEN, 2-1, Hirosawa, Wako-shi, Saitama 351-0198 (JP). **HARBERS, Matthias, T.** [DE/JP]; c/o KABUSHIKI KAISHA

Published:

— with international search report

[Continued on next page]

(54) Title: METHOD OF UTILIZING THE 5'END OF TRANSCRIBED NUCLEIC ACID REGIONS FOR CLONING AND ANALYSIS



(57) Abstract: A method is disclosed for obtaining the 5' ends of transcribed regions from a plurality of nucleic acid fragments obtained from biological materials or synthetic pools. DNA fragments encoding the 5' ends are enriched for their individual analysis or for the analysis of concatemers thereof. The sequence information derived from 5' ends can be used for characterization and cloning of the transcriptome.

WO 2003/106672 A3



(88) Date of publication of the international search report:
1 July 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/106672

SEQUENCE LISTING

<110> RIKEN
 KABUSHIKI KAISHA DNAFORM
 <120> Method for utilizing the 5' end of mRNA for cloning and analysis
 <130> 1336(PCT)
 <150> JP 2002-171851
 <151> 2002-06-12
 <150> JP 2002-235294
 <151> 2002-08-12
 <160> 77
 <170> PatentIn version 3.1
 <210> 1
 <211> 74
 <212> DNA
 <213> Artificial
 <220>
 <223> First strand cDNA primer
 <220>
 <221> misc_feature
 <222> (73)..(73)
 <223> "v" is A, C or G
 <220>
 <221> misc_feature
 <222> (74)..(74)
 <223> "n" is any nucleotide
 <400> 1
 gagagagaga aaggatcctg ccatttcatt acctctttct ccgcacccga catagatttt 60
 tttttttttt ttvn 74
 <210> 2
 <211> 60
 <212> DNA
 <213> Artificial
 <220>
 <223> Upper oligonucleotide GN5
 <220>
 <221> misc_feature
 <222> (56)..(60)
 <223> "n" is any nucleotide
 <400> 2
 agagagagac ctcgagtaac tataacgggc ctaaggtagc gacctaggtc cgacgnnnnn 60
 <210> 3
 <211> 60
 <212> DNA
 <213> Artificial
 <220>
 <223> Upper oligonucleotide N6

<220>
<221> misc_feature
<222> (55)..(60)
<223> "n" is any nucleotide

<400> 3
agagagagac ctcgagtaac tataacggtc ctaaggtagc gacctaggtc cgacnnnnnn 60

<210> 4
<211> 54
<212> DNA
<213> Artificial

<220>
<223> Lower oligonucleotide

<400> 4
gtcggaccta ggtcgctacc ttaggaccgt tatagttact cgaggctctt ctct 54

<210> 5
<211> 55
<212> DNA
<213> Artificial

<220>
<223> primer

<400> 5
agagagagac ctcgagtaac tataacggtc ctaaggtagc gacctaggtc cgacg 55

<210> 6
<211> 45
<212> DNA
<213> Artificial

<220>
<223> linker

<400> 6
tctagatcag gactcttcta tagtgcacc taaagtctct ctctc 45

<210> 7
<211> 47
<212> DNA
<213> Artificial

<220>
<223> linker

<220>
<221> misc_feature
<222> (46)..(47)
<223> "n" is any nucleotide

<400> 7
gagagagaga ctttaggtga cactatagaa ggtcctgat ctagann 47

<210> 8
<211> 25
<212> DNA
<213> Artificial

<220>
 <223> Primer 1 (uni-PCR)
 <400> 8
 gagagagaga ctttaggtga cacta 25
 <210> 9
 <211> 25
 <212> DNA
 <213> Artificial
 <220>
 <223> Primer 2(MneI-PCR)
 <400> 9
 agagagagac ctcgagtaac tataa 25
 <210> 10
 <211> 44
 <212> DNA
 <213> Artificial
 <220>
 <223> first strand oligo-dT primer
 <220>
 <221> misc_feature
 <222> (43)..(43)
 <223> "v" is A, C or G
 <220>
 <221> misc_feature
 <222> (44)..(44)
 <223> "n" is any nucleotide
 <400> 10
 gagagagaga ggatccttct ggagagtttt tttttttttt ttvn 44
 <210> 11
 <211> 45
 <212> DNA
 <213> Artificial
 <220>
 <223> Oligonucleotide Bg-Gsu-GN5
 <220>
 <221> misc_feature
 <222> (41)..(45)
 <223> "n" is any nucleotide
 <400> 11
 agagagagaa ctaggcttaa taggtgacta gatctggagg nnnnn 45
 <210> 12
 <211> 45
 <212> DNA
 <213> Artificial
 <220>
 <223> Oligonucleotide Bg-Gsu-N6
 <220>

<221> misc_feature
 <222> (40)..(45)
 <223> "n" is any nucleotide

<400> 12
 agagagagaa ctaggcttaa taggtgacta gatctggagn nnnnn 45

<210> 13
 <211> 39
 <212> DNA
 <213> Artificial

<220>
 <223> Oligonucleotide Bg-Gsu-down

<400> 13
 ctggagatct agtcacctat taagcctagt tctctctct 39

<210> 14
 <211> 47
 <212> DNA
 <213> Artificial

<220>
 <223> Oligonucleotide Bg-Mme-GN5

<220>
 <221> misc_feature
 <222> (43)..(47)
 <223> "n" is any nucleotide

<220>
 <221> misc_feature
 <222> (39)..(39)
 <223> "r" is G or A

<400> 14
 agagagagaa ctaggcttaa taggtgacta gatcttcra cgnnnnn 47

<210> 15
 <211> 47
 <212> DNA
 <213> Artificial

<220>
 <223> Oligonucleotide Bg-Mme-N6

<220>
 <221> misc_feature
 <222> (42)..(47)
 <223> "n" is any nucleotide

<220>
 <221> misc_feature
 <222> (39)..(39)
 <223> "r" is G or A

<400> 15
 agagagagaa ctaggcttaa taggtgacta gatcttcra cnnnnnn 47

<210> 16

<211> 40
 <212> DNA
 <213> Artificial

<220>
 <223> Oligonucleotide Bg-Mme-down

<220>
 <221> misc_feature
 <222> (3)..(3)
 <223> "y" is C or T

<400> 16
 gtyggagatc tagtcaccta ttaagcctag ttctctctct

40

<210> 17
 <211> 47
 <212> DNA
 <213> Artificial

<220>
 <223> oligonucleotide

<220>
 <221> misc_feature
 <222> (46)..(47)
 <223> "n" is any nucleotide

<400> 17
 gagagagaga ctttaggtga cactatagaa gagtcctgag aattcnn

47

<210> 18
 <211> 45
 <212> DNA
 <213> Artificial

<220>
 <223> oligonucleotide

<400> 18
 gaattctcag gactcttcta tagtgtcacc taaagtctct ctctc

45

<210> 19
 <211> 49
 <212> DNA
 <213> Artificial

<220>
 <223> linker

<220>
 <221> misc_feature
 <222> (45)..(49)
 <223> "n" is any nucleotide

<400> 19
 agagagagag cttagatgag agtgactcga gcctaggtcc aacgnnnnn

49

<210> 20
 <211> 49
 <212> DNA
 <213> Artificial

<220>

<223> linker

<220>

<221> misc_feature

<222> (44)..(49)

<223> "n" is any nucleotide

<400> 20

agagagagag cttagatgag agtgactcga gcctaggtcc aacnnnnnn

49

<210> 21

<211> 43

<212> DNA

<213> Artificial

<220>

<223> linker

<400> 21

gttggaccta ggctcgagtc actctcatct aagctctctc tct

43

<210> 22

<211> 42

<212> DNA

<213> Artificial

<220>

<223> second linker

<400> 22

gaattctacg cctctcgatc gaaatcccg tctaggctag cg

42

<210> 23

<211> 42

<212> DNA

<213> Artificial

<220>

<223> second linker

<400> 23

cttaagatgc ggagagcgtg aatcgagttt aaggctagca tc

42

<210> 24

<211> 25

<212> DNA

<213> Artificial

<220>

<223> primer

<400> 24

ttagatgaga gtgactcgag cctag

25

<210> 25

<211> 25

<212> DNA

<213> Artificial

<220>

<223> primer

<400> 25

ctacgatcgg aatttgagct aagtg

25

<210> 26
 <211> 17
 <212> DNA
 <213> Artificial

<220>
 <223> primer

<400> 26
 caggaaacag ctatgac

17

<210> 27
 <211> 16
 <212> DNA
 <213> Artificial

<220>
 <223> primer

<400> 27
 gtaaaacgac ggccag

16

<210> 28
 <211> 596
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (432)..(432)
 <223> "n" is any nucleotide

<220>
 <221> misc_feature
 <222> (436)..(436)
 <223> "n" is any nucleotide

<220>
 <221> misc_feature
 <222> (456)..(456)
 <223> "n" is any nucleotide

<400> 28
 tcgttaacta ttaggcgaat tgggccctct aggtcgacga gttctcagca gagccgccgt 60
 ctagagcccc gccctcccgg gccaccgtcg gacctagaat agttactcga ggtctctcgt 120
 cggacctaga gtttttcgta tgtttgtcat cgtcggacct aggtccgacg gtccattcct 180
 gagagtctct ctaggtccga cgagagagag aggatccttc tgtctagacc ctgacgccgg 240
 aaccgcaccg tcggacctag gtccgacgga aaagcagctt cctccactct aggtccgacg 300
 gtgtgtgtgt gtgtgcgtgt tctagagact ggttcagatc aaaagtcgtc ggacctaggt 360
 ccgacggggc tggtagatg gctcagtcta gatgcatgct cgagcggccg ccagtgtgat 420
 ggatatctgc cnaatnccag cacaccggcg cgcgcnacca gtggatccga gcccggtacc 480
 aagcttgatg catacctcga gtatcctata ctgtcaccta aatagcttgg ggtaatcatg 540
 gtcatagctg tctcctgtgt gaaattgita tccgctcaaa attccaaca acatag 596

<210> 29
 <211> 12
 <212> DNA
 <213> Artificial

<220>
 <223> linker

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> "n" is any nucleotide

<400> 29
 nctaggtccg ac

12

<210> 30
 <211> 20
 <212> DNA
 <213> Artificial

<220>
 <223> linker

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> "n" is any nucleotide

<220>
 <221> misc_feature
 <222> (20)..(20)
 <223> "n" is any nucleotide

<400> 30
 ngttggacct aggtccaacn

20

<210> 31
 <211> 13
 <212> DNA
 <213> Artificial

<220>
 <223> linker

<400> 31
 tctaggtccg acg

13

<210> 32
 <211> 13
 <212> DNA
 <213> Artificial

<220>
 <223> linker

<400> 32
 cctaggtccg acg

13

<210> 33
 <211> 20

<212> DNA
<213> Artificial

<220>
<223> tag1

<400> 33
gtggcccggg agggcggggc 20

<210> 34
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag2

<400> 34
agagacctcg agtaactat 19

<210> 35
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag3

<400> 35
atgacaaaca tacgaaaac 20

<210> 36
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag4

<400> 36
gtccattcct gagagtctc 19

<210> 37
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag5

<400> 37
agagagagag gatccttctg 20

<210> 38
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag6

<400> 38
gtgcggttcc ggcgtcaggg 20

<210> 39
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag7

<400> 39
gaaaagcagc ttcctccac 19

<210> 40
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag8

<400> 40
gtgtgtgtgt gtgtgcgtgt 20

<210> 41
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag9

<400> 41
acttttgatc tgaaccagtc 20

<210> 42
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag10

<400> 42
gggctggtga gatggctcag 20

<210> 43
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag1

<400> 43
gtacctctc gcatcccg 19

<210> 44
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag2

<400> 44
gtggtgtgcg tgcgaagg 20

<210> 45
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag3

<400> 45
atggaccgag ggcggcagcc

20

<210> 46
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag4

<400> 46
cggatcgggt gggtcggac

19

<210> 47
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag5

<400> 47
agaggtcgca gcagttcgt

19

<210> 48
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag6

<400> 48
tctccggagc cggcgtgtg

20

<210> 49
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag7

<400> 49
agactttgca ggctccgag

19

<210> 50
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag8

<400> 50
ggagctgccg cagcgccgga 20

<210> 51
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag9

<400> 51
acaccgtcgg acctggtcgc 20

<210> 52
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag10

<400> 52
agacgttctc gccagagtc 20

<210> 53
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag11

<400> 53
gccgttcctt gcttgctgga 20

<210> 54
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag12

<400> 54
gggttgggga tttagctcag 20

<210> 55
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag13

<400> 55
gagtaactat aacggtcct 19

<210> 56
<211> 19
<212> DNA
<213> Artificial

<220>

<223> tag14

<400> 56
gattccgcct ggagctcgc

19

<210> 57
<211> 18
<212> DNA
<213> Artificial

<220>
<223> tag15

<400> 57
agggaccgct gcggtccg

18

<210> 58
<211> 18
<212> DNA
<213> Artificial

<220>
<223> zzb21106i09t3 junk

<400> 58
cattagggga ttggccc

18

<210> 59
<211> 28
<212> DNA
<213> Artificial

<220>
<223> zzb21106i09t3 junk

<400> 59
acccgggggg cgggactaac cgtcggac

28

<210> 60
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag1

<400> 60
gacgcggaag gcgcggcggc

20

<210> 61
<211> 21
<212> DNA
<213> Artificial

<220>
<223> tag2

<400> 61
ggagggcggg cggcgccct c

21

<210> 62
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag3

<400> 62
caaaaaaaaaa aaaaaaact 19

<210> 63
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag4

<400> 63
aggctctgct cgctctgcc 20

<210> 64
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag5

<400> 64
acttctgatt ctgacagac 19

<210> 65
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag6

<400> 65
acagtggcgt ctgcaaagc 19

<210> 66
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag7

<400> 66
ggaaagtcca ggtggacttt 20

<210> 67
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag8

<400> 67
gccgccgagg ccgcgagg 19

<210> 68
<211> 19

<212> DNA
<213> Artificial

<220>
<223> tag9

<400> 68
gttagtgtat aacagagtt 19

<210> 69
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag10

<400> 69
tcgcccgtg ttcagtctct 20

<210> 70
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag11

<400> 70
aggtggggca agatggctg 19

<210> 71
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag12

<400> 71
ggcatggcca gaaggcaagc 20

<210> 72
<211> 20
<212> DNA
<213> Artificial

<220>
<223> tag13

<400> 72
gacgcacgca tagagggggg 20

<210> 73
<211> 19
<212> DNA
<213> Artificial

<220>
<223> tag14

<220>
<221> misc_feature
<222> (1)..(1)
<223> "n" is any nucleotide

<400> 73
 nccatggaac agccacact 19

<210> 74
 <211> 26
 <212> DNA
 <213> Artificial

<220>
 <223> junk1

<400> 74
 tgataaggca atggcctcta atgctg 26

<210> 75
 <211> 16
 <212> DNA
 <213> Artificial

<220>
 <223> tag

<400> 75
 acctccctcc gcggag 16

<210> 76
 <211> 2745
 <212> DNA
 <213> Homo sapiens

<220>
 <221> CDS
 <222> (842)..(2017)
 <223>

<400> 76
 acctccctcc gcggagcagc cagacagcga gggccccggc cgggggcagg ggggacgccc 60
 cgtccggggc accccccccg gctctgagcc gcccgcgggg ccggcctcgg ccgggagcgg 120
 aggaaggagt cgccgaggag cagcctgagg ccccagagtc tgagacgagc cgccgcccgc 180
 cccgccactg cggggaggag ggggaggagg agcgggagga gggacgagct ggtcgggaga 240
 agaggaaaaa aactttttag acttttccgt tgccgctggg agccggaggc gcggggacct 300
 cttggcgcca cgctgccccg cgaggaggca ggacttgggg accccagacc gcctcccttt 360
 gccgccgggg acgtttgctc cctccctgcc ccctacacgg cgtccctcag gcgccccat 420
 tccggaccag ccctcgggag tcgccgaccc ggcctcccgc aaagactttt ccccagacct 480
 cgggcgacc ccctgcacgc cgccttcac cccggcctgt ctctgagcc ccgcgcac 540
 ctagaccctt tctcctccag gagacggatc tctctccgac ctgccacaga tcccctattc 600
 aagaccaccc accttctggt accagatcgc gcccatctag gttatttcg tgggatactg 660
 agacaccccc ggtccaagcc tcccctccac cactgcgccc ttctccctga ggagcctcag 720
 ctttccctcg aggcctcct accttttgc gggagacccc cagcccctgc aggggcgggg 780
 cctccccacc acaccagccc tgttcgcgt ctcggcagtg ccggggggcg ccgcctcccc 840
 c atg ccg ccc tcc ggg ctg cgg ctg ctg ccg ctg ctg cta ccg ctg ctg 889

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|--|
| Met | Pro | Pro | Ser | Gly | Leu | Arg | Leu | Leu | Pro | Leu | Leu | Leu | Pro | Leu | Leu | | |
| 1 | | | | 5 | | | | | 10 | | | | | | 15 | | |
| tgg | cta | ctg | gtg | ctg | acg | cct | ggc | ccg | ccg | gcc | gcg | gga | cta | tcc | acc | 937 | |
| Trp | Leu | Leu | Val | Leu | Thr | Pro | Gly | Pro | Pro | Ala | Ala | Gly | Leu | Ser | Thr | | |
| | | | 20 | | | | | 25 | | | | | 30 | | | | |
| tgc | aag | act | atc | gac | atg | gag | ctg | gtg | aag | cgg | aag | cgc | atc | gag | gcc | 985 | |
| Cys | Lys | Thr | Ile | Asp | Met | Glu | Leu | Val | Lys | Arg | Lys | Arg | Ile | Glu | Ala | | |
| | | 35 | | | | 40 | | | | | | 45 | | | | | |
| atc | cgc | ggc | cag | atc | ctg | tcc | aag | ctg | cgg | ctc | gcc | agc | ccc | ccg | agc | 1033 | |
| Ile | Arg | Gly | Gln | Ile | Leu | Ser | Lys | Leu | Arg | Leu | Ala | Ser | Pro | Pro | Ser | | |
| | 50 | | | | | 55 | | | | | 60 | | | | | | |
| cag | ggg | gag | gtg | ccg | ccc | ggc | ccg | ctg | ccc | gag | gcc | gtg | ctc | gcc | ctg | 1081 | |
| Gln | Gly | Glu | Val | Pro | Pro | Gly | Pro | Leu | Pro | Glu | Ala | Val | Leu | Ala | Leu | | |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 | | |
| tac | aac | agc | acc | cgc | gac | cgg | gtg | gcc | ggg | gag | agt | gca | gaa | ccg | gag | 1129 | |
| Tyr | Asn | Ser | Thr | Arg | Asp | Arg | Val | Ala | Gly | Glu | Ser | Ala | Glu | Pro | Glu | | |
| | | | | 85 | | | | | 90 | | | | | 95 | | | |
| ccc | gag | cct | gag | gcc | gac | tac | tac | gcc | aag | gag | gtc | acc | cgc | gtg | cta | 1177 | |
| Pro | Glu | Pro | Glu | Ala | Asp | Tyr | Tyr | Ala | Lys | Glu | Val | Thr | Arg | Val | Leu | | |
| | | | 100 | | | | | 105 | | | | | 110 | | | | |
| atg | gtg | gaa | acc | cac | aac | gaa | atc | tat | gac | aag | ttc | aag | cag | agt | aca | 1225 | |
| Met | Val | Glu | Thr | His | Asn | Glu | Ile | Tyr | Asp | Lys | Phe | Lys | Gln | Ser | Thr | | |
| | | 115 | | | | 120 | | | | | | 125 | | | | | |
| cac | agc | ata | tat | atg | ttc | ttc | aac | aca | tca | gag | ctc | cga | gaa | gcg | gta | 1273 | |
| His | Ser | Ile | Tyr | Met | Phe | Phe | Asn | Thr | Ser | Glu | Leu | Arg | Glu | Ala | Val | | |
| | 130 | | | | | 135 | | | | | | 140 | | | | | |
| cct | gaa | ccc | gtg | ttg | ctc | tcc | cgg | gca | gag | ctg | cgt | ctg | ctg | agg | agg | 1321 | |
| Pro | Glu | Pro | Val | Leu | Leu | Ser | Arg | Ala | Glu | Leu | Arg | Leu | Leu | Arg | Arg | | |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 | | |
| ctc | aag | tta | aaa | gtg | gag | cag | cac | gtg | gag | ctg | tac | cag | aaa | tac | agc | 1369 | |
| Leu | Lys | Leu | Lys | Val | Glu | Gln | His | Val | Glu | Leu | Tyr | Gln | Lys | Tyr | Ser | | |
| | | | | 165 | | | | | 170 | | | | | 175 | | | |
| aac | aat | tcc | tgg | cga | tac | ctc | agc | aac | cgg | ctg | ctg | gca | ccc | agc | gac | 1417 | |
| Asn | Asn | Ser | Trp | Arg | Tyr | Leu | Ser | Asn | Arg | Leu | Leu | Ala | Pro | Ser | Asp | | |
| | | | 180 | | | | | 185 | | | | | 190 | | | | |
| tcg | cca | gag | tgg | tta | tct | ttt | gat | gtc | acc | gga | gtt | gtg | cgg | cag | tgg | 1465 | |
| Ser | Pro | Glu | Trp | Leu | Ser | Phe | Asp | Val | Thr | Gly | Val | Val | Arg | Gln | Trp | | |
| | | 195 | | | | | 200 | | | | | 205 | | | | | |
| ttg | agc | cgt | gga | ggg | gaa | att | gag | ggc | ttt | cgc | ctt | agc | gcc | cac | tgc | 1513 | |
| Leu | Ser | Arg | Gly | Gly | Glu | Ile | Glu | Gly | Phe | Arg | Leu | Ser | Ala | His | Cys | | |
| | 210 | | | | | 215 | | | | | 220 | | | | | | |
| tcc | tgt | gac | agc | agg | gat | aac | aca | ctg | caa | gtg | gac | atc | aac | ggg | ttc | 1561 | |
| Ser | Cys | Asp | Ser | Arg | Asp | Asn | Thr | Leu | Gln | Val | Asp | Ile | Asn | Gly | Phe | | |
| | 225 | | | | | 230 | | | | 235 | | | | | 240 | | |
| act | acc | ggc | cgc | cga | ggt | gac | ctg | gcc | acc | att | cat | ggc | atg | aac | cgg | 1609 | |
| Thr | Thr | Gly | Arg | Arg | Gly | Asp | Leu | Ala | Thr | Ile | His | Gly | Met | Asn | Arg | | |
| | | | | 245 | | | | | 250 | | | | | 255 | | | |
| cct | ttc | ctg | ctt | ctc | atg | gcc | acc | cgg | ctg | gag | agg | gcc | cag | cat | ctg | 1657 | |
| Pro | Phe | Leu | Leu | Leu | Met | Ala | Thr | Pro | Leu | Glu | Arg | Ala | Gln | His | Leu | | |
| | | | 260 | | | | | 265 | | | | | | 270 | | | |
| caa | agc | tcc | cgg | cac | cgc | cga | gcc | ctg | gac | acc | aac | tat | tgc | ttc | agc | 1705 | |
| Gln | Ser | Ser | Arg | His | Arg | Arg | Ala | Leu | Asp | Thr | Asn | Tyr | Cys | Phe | Ser | | |
| | | 275 | | | | | 280 | | | | | | 285 | | | | |

tcc acg gag aag aac tgc tgc gtg cgg cag ctg tac att gac ttc cgc 1753
 Ser Thr Glu Lys Asn Cys Cys Val Arg Gln Leu Tyr Ile Asp Phe Arg
 290 295 300

aag gac ctc ggc tgg aag tgg atc cac gag ccc aag ggc tac cat gcc 1801
 Lys Asp Leu Gly Trp Lys Trp Ile His Glu Pro Lys Gly Tyr His Ala
 305 310 315 320

aac ttc tgc ctc ggg ccc tgc ccc tac att tgg agc ctg gac acg cag 1849
 Asn Phe Cys Leu Gly Pro Cys Pro Tyr Ile Trp Ser Leu Asp Thr Gln
 325 330 335

tac agc aag gtc ctg gcc ctg tac aac cag cat aac ccg ggc gcc tcg 1897
 Tyr Ser Lys Val Leu Ala Leu Tyr Asn Gln His Asn Pro Gly Ala Ser
 340 345 350

gcg gcg ccg tgc tgc gtg ccg cag gcg ctg gag ccg ctg ccc atc gtg 1945
 Ala Ala Pro Cys Cys Val Pro Gln Ala Leu Glu Pro Leu Pro Ile Val
 355 360 365

tac tac gtg ggc cgc aag ccc aag gtg gag cag ctg tcc aac atg atc 1993
 Tyr Tyr Val Gly Arg Lys Pro Lys Val Glu Gln Leu Ser Asn Met Ile
 370 375 380

gtg cgc tcc tgc aag tgc agc tga ggtcccgccc cgccccgccc cgccccggca 2047
 Val Arg Ser Cys Lys Cys Ser
 385 390

ggccccggccc caccgccccc cgcccccgct gccttgccca tgggggctgt atttaaggac 2107

accgtgcccc aagccacac ggggccccat taaagatgga gagaggactg cggatctctg 2167

tgctattggg cgctgcctg gggctctccat cctgacgtt cccccactcc cactccctct 2227

ctctccctct ctgcctctc ctgcctgtct gcactattcc tttgccggc atcaaggcac 2287

aggggaccag tggggaacac tactgtagtt agatctattt attgagcacc ttgggcactg 2347

ttgaagtgcc ttacattaat gaactcattc agtcaccata gcaacactct gagatggcag 2407

ggactctgat aacaccacatt ttaaagggtg aggaacaac ccagagagg ttaaggagg 2467

agttcctgcc caccaggaac ctgctttagt gggggatagt gaagaagaca ataaaagata 2527

gtagttcagg ccaggcgggg tgctcacgcc tgtaatccta gcacttttgg gaggcagaga 2587

tgggaggata cttgaatcca ggcatttgag accagcctgg gtaacatagt gagaccctat 2647

ctctacaaaa cacttttaaa aaatgtacac ctgtggtccc agctactctg gaggctaagg 2707

tgggaggatc acttgatcct gggaggtcaa ggctgcag 2745

<210> 77
 <211> 391
 <212> PRT
 <213> Homo sapiens

<400> 77

Met Pro Pro Ser Gly Leu Arg Leu Leu Pro Leu Leu Leu Pro Leu Leu
 1 5 10 15

Trp Leu Leu Val Leu Thr Pro Gly Pro Pro Ala Ala Gly Leu Ser Thr
 20 25 30

Cys Lys Thr Ile Asp Met Glu Leu Val Lys Arg Lys Arg Ile Glu Ala

| 35 | 40 | 45 |
|--|----|----|
| Ile Arg Gly Gln Ile Leu Ser Lys Leu Arg Leu Ala Ser Pro Pro Ser 50 55 60 | | |
| Gln Gly Glu Val Pro Pro Gly Pro Leu Pro Glu Ala Val Leu Ala Leu 65 70 75 80 | | |
| Tyr Asn Ser Thr Arg Asp Arg Val Ala Gly Glu Ser Ala Glu Pro Glu 85 90 95 | | |
| Pro Glu Pro Glu Ala Asp Tyr Tyr Ala Lys Glu Val Thr Arg Val Leu 100 105 110 | | |
| Met Val Glu Thr His Asn Glu Ile Tyr Asp Lys Phe Lys Gln Ser Thr 115 120 125 | | |
| His Ser Ile Tyr Met Phe Phe Asn Thr Ser Glu Leu Arg Glu Ala Val 130 135 140 | | |
| Pro Glu Pro Val Leu Leu Ser Arg Ala Glu Leu Arg Leu Leu Arg Arg 145 150 155 160 | | |
| Leu Lys Leu Lys Val Glu Gln His Val Glu Leu Tyr Gln Lys Tyr Ser 165 170 175 | | |
| Asn Asn Ser Trp Arg Tyr Leu Ser Asn Arg Leu Leu Ala Pro Ser Asp 180 185 190 | | |
| Ser Pro Glu Trp Leu Ser Phe Asp Val Thr Gly Val Val Arg Gln Trp 195 200 205 | | |
| Leu Ser Arg Gly Gly Glu Ile Glu Gly Phe Arg Leu Ser Ala His Cys 210 215 220 | | |
| Ser Cys Asp Ser Arg Asp Asn Thr Leu Gln Val Asp Ile Asn Gly Phe 225 230 235 240 | | |
| Thr Thr Gly Arg Arg Gly Asp Leu Ala Thr Ile His Gly Met Asn Arg 245 250 255 | | |
| Pro Phe Leu Leu Leu Met Ala Thr Pro Leu Glu Arg Ala Gln His Leu 260 265 270 | | |
| Gln Ser Ser Arg His Arg Arg Ala Leu Asp Thr Asn Tyr Cys Phe Ser 275 280 285 | | |
| Ser Thr Glu Lys Asn Cys Cys Val Arg Gln Leu Tyr Ile Asp Phe Arg 290 295 300 | | |
| Lys Asp Leu Gly Trp Lys Trp Ile His Glu Pro Lys Gly Tyr His Ala 305 310 315 320 | | |

Asn Phe Cys Leu Gly Pro Cys Pro Tyr Ile Trp Ser Leu Asp Thr Gln
325 330 335

Tyr Ser Lys Val Leu Ala Leu Tyr Asn Gln His Asn Pro Gly Ala Ser
340 345 350

Ala Ala Pro Cys Cys Val Pro Gln Ala Leu Glu Pro Leu Pro Ile Val
355 360 365

Tyr Tyr Val Gly Arg Lys Pro Lys Val Glu Gln Leu Ser Asn Met Ile
370 375 380

Val Arg Ser Cys Lys Cys Ser
385 390

INTERNATIONAL SEARCH REPORT

PCT/JP 03/07514

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category ° | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------------------|
| X | <p>WO 02 10438 A (JOHNS HOPKINS UNIVERSITY) 7 February 2002 (2002-02-07) cited in the application the whole document see especially: page 4, paragraphs 12,13 page 10, paragraph 31 page 12, paragraph 34 page 14, paragraph 39 -page 15, paragraph 41 page 16, paragraph 43 page 19, paragraph 54 -page 20, paragraph 56 page 25 -page 42; examples 1-5 page 43 -page 60; claims 1-64,75-113 page 61 -page 62; figure 1</p> <p style="text-align: center;">-/-</p> | <p>1-6, 8-16, 26-55</p> |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

25 February 2004

Date of mailing of the international search report

03/03/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Fuchs, U

INTERNATIONAL SEARCH REPORT

PCT/JP 03/07514

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|-----------------------|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | MARUYAMA, K. & SUGANO, S.: "Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides" GENE, vol. 138, no. 1-2, 28 January 1994 (1994-01-28), pages 171-174, XP001008855 cited in the application the whole document --- | 1-55 |
| A | SHIBATA, Y. ET AL.: "Cloning Full-Length, Cap-Trapper-Selected cDNAs by Using the Single-Strand Linker Ligation Method" BIOTECHNIQUES, vol. 30, no. 6, June 2001 (2001-06), pages 1250-1254, XP002197302 cited in the application the whole document --- | 1-55 |
| A | SUZUKI, Y. ET AL.: "Diverse transcriptional initiation revealed by fine, large-scale mapping of mRNA start sites" EMBO REPORTS, vol. 2, no. 5, May 2001 (2001-05), pages 388-393, XP001176665 cited in the application the whole document --- | 1-55 |
| A | CARNINCI, P. & HAYASHIZAKI, Y.: "High-Efficiency Full-Length cDNA Cloning" METHODS IN ENZYMOLOGY, vol. 303, 1999, pages 19-44, XP001022979 cited in the application the whole document --- | 1-55 |
| A | EDERY, I. ET AL.: "An Efficient Strategy To Isolate Full-Length cDNAs Based on an mRNA Cap Retention Procedure (CAPture)" MOLECULAR AND CELLULAR BIOLOGY, vol. 15, no. 6, June 1995 (1995-06), pages 3363-3371, XP002915963 cited in the application the whole document --- | 1-55 |
| | --- | |

-/-

INTERNATIONAL SEARCH REPORT

PCT/JP 03/07514

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT | | |
|--|--|---|
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| A | THEISSEN, H. ET AL.: "Cloning of the human cDNA for the U1 RNA-associated 70K protein" EMBO JOURNAL, vol. 5, no. 12, 1 December 1986 (1986-12-01), pages 3209-3217, XP009002534 cited in the application page 3211, column 1, line 38 -column 2, line 16 page 3212 -page 3213; figure 3 | 1-55 |
| A | EP 1 197 552 A (RIKEN; HAYASHIZAKI, Y.) 17 April 2002 (2002-04-17) cited in the application the whole document | 1-55 |
| A | US 6 352 828 B1 (BRENNER, S.) 5 March 2002 (2002-03-05) cited in the application the whole document | 1-55 |
| E | WO 03 091416 A (LYNX THERAPEUTICS, INC.) 6 November 2003 (2003-11-06) the whole document see especially: page 18, line 25 -page 21, line 18 and page 41 -page 43; figures 3A-C | 1-8, 11, 12, 14-16, 26-30, 32, 37-45, 48-55 |
| T | SHIRAKI, T. ET AL.: "Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 100, no. 26, 23 December 2003 (2003-12-23), pages 15776-15781, XP001161070 the whole document | 1-55 |

INTERNATIONAL SEARCH REPORT

International Application No

PCT/JP 03/07514

| Patent document cited in search report | | Publication date | Patent family member(s) | Publication date |
|---|----|---------------------|----------------------------|---------------------|
| WO 0210438 | A | 07-02-2002 | AU 8087501 A | 13-02-2002 |
| | | | WO 0210438 A2 | 07-02-2002 |
| | | | US 2003008290 A1 | 09-01-2003 |
| EP 1197552 | A | 17-04-2002 | CA 2356123 A1 | 25-02-2002 |
| | | | EP 1197552 A2 | 17-04-2002 |
| | | | JP 2002253237 A | 10-09-2002 |
| | | | US 2002106666 A1 | 08-08-2002 |
| US 6352828 | B1 | 05-03-2002 | US 5846719 A | 08-12-1998 |
| | | | US 5604097 A | 18-02-1997 |
| | | | AU 733782 B2 | 24-05-2001 |
| | | | AU 3374097 A | 05-01-1998 |
| | | | CA 2256700 A1 | 11-12-1997 |
| | | | CN 1230226 A | 29-09-1999 |
| | | | CZ 9803979 A3 | 14-07-1999 |
| | | | EP 0923650 A1 | 23-06-1999 |
| | | | HU 0003944 A2 | 28-03-2001 |
| | | | JP 2000515006 T | 14-11-2000 |
| | | | NO 985698 A | 08-02-1999 |
| | | | PL 331513 A1 | 19-07-1999 |
| | | | US 6138077 A | 24-10-2000 |
| | | | US 6172218 B1 | 09-01-2001 |
| | | | WO 9746704 A1 | 11-12-1997 |
| | | | US 6235475 B1 | 22-05-2001 |
| | | | US 6172214 B1 | 09-01-2001 |
| | | | US 6150516 A | 21-11-2000 |
| | | | US 6013445 A | 11-01-2000 |
| | | | AU 712929 B2 | 18-11-1999 |
| | | | AU 4277896 A | 06-05-1996 |
| | | | AU 5266399 A | 09-12-1999 |
| | | | CA 2202167 A1 | 25-04-1996 |
| | | | CA 2441634 A1 | 25-04-1996 |
| | | | CZ 9700866 A3 | 17-09-1997 |
| | | | EP 0793718 A1 | 10-09-1997 |
| | | | FI 971473 A | 04-06-1997 |
| | | | HU 77916 A2 | 28-10-1998 |
| | | | JP 10507357 T | 21-07-1998 |
| | | | NO 971644 A | 02-06-1997 |
| | | | US 6280935 B1 | 28-08-2001 |
| | | | WO 9612014 A1 | 25-04-1996 |
| | | | US 5635400 A | 03-06-1997 |
| | | | US 5654413 A | 05-08-1997 |
| | | | AU 3946195 A | 06-05-1996 |
| | | | DE 69513997 D1 | 20-01-2000 |
| | | | DE 69513997 T2 | 27-07-2000 |
| | | | EP 0786014 A1 | 30-07-1997 |
| | | | EP 0952216 A2 | 27-10-1999 |
| | | | WO 9612039 A1 | 25-04-1996 |
| | | | US 6140489 A | 31-10-2000 |
| | | | US 5695934 A | 09-12-1997 |
| | | | US 5863722 A | 26-01-1999 |
| WO 03091416 | A | 06-11-2003 | WO 03091416 A2 | 06-11-2003 |
| | | | US 2004002104 A1 | 01-01-2004 |